DSC291 Spring 2022
Large Scale Statistical Analysis

Project guidelines

The purpose of the project is to give you the opportunity to apply the concepts learned in class to a problem that is of interest to you. The problem should center around multiple regression with a continuous outcome. The data analysis should be as thorough as possible using the concepts learned in class.

Project proposal (due 05 May 2022):
Report: (~3 pages) As a guideline, the project proposal report may include the following items:
1. A title for your project.
2. A brief summary of the scientific background, what variables are being measured and their type (continuous, categorical, ordinal, etc.) and what are the questions of interest (~1/2 page).
3. A (full or partial) plot (or table) of the data.
4. Qualitative description of the data: sampling size, basic summary statistics, apparent trends, etc.
5. Initial analysis: multiple regression, initial assessment and interpretation, basic diagnostics.

Presentation: A ~10 min. presentation summarizing the report, 5 slides (about one slide per item above).

Final report (due 02 June 2022):
Report: (~10 pages) As a guideline, the project final report may include the following items:
1. Background and objective as in the project proposal, but sharper and fuller (including references).
2. Relevant theoretical calculations, if used.
3. Full data analysis, including diagnostics, model selection, etc. A clear explanation of the data analysis and the thought process accompanying it.
4. Plots, results.
5. Conclusions - were the objectives met?
6. Computer code.

Presentation: A ~20 min. presentation summarizing the report, 10 slides (about 2 min. per slide).

Grading rubrics

Proposal:

- Title
- Contribution statement
- Scientific background
- Problem description
- Data description
- Analysis plan
- Writing quality and length

Proposal presentation:

- Graphics/text composition
- Speaking
- Confidence/knowledge
- Timing
- Participation in the audience